

U.S. Department of the Interior
U.S. Geological Survey

Registering Web Map Services in *The National Map Catalog*

Version 0.9.1

Mid-Continent Mapping Center
March 2005

This document is altered frequently. The current version can always be found at <http://thor-f5.er.usgs.gov/nmcatalog/>

This document is for use by USGS production personnel, data partners, and contractors. It has not been reviewed for conformance with USGS editorial standards and has not been approved for formal publication. Any use of trade names is for descriptive purposes only and does not constitute endorsement by the US Government.

Release Notes and Document History

Version 0.4.0, August 6, 2003. First version to be released for review throughout the Geography Discipline. Earlier versions were for technical review within MCMC. Original title was "Technical Aspects of *The National Map* Data Partnerships."

Version 0.5.0, September 12, 2003. Incorporates comments from the first Discipline-wide review. Cross-references to other *The National Map* documents added; the metadata policy question addressed more explicitly; clarifications made to the document scope, particularly the fact that data production procedures of USGS national data programs are not within the scope of the document.

Version 0.6.0, October 2003. Attachment C added: discussion of policy and procedures for implementing a partnership, populating the catalog, and making new data visible in the public viewer.

Version 0.7.0, November 2003. New section added to provide more of a cookbook description of information needed in an agreement. More information on min/max viewscales added to Attachment B. General editing to incorporate comments and bring the document up to date with recent Steering Team guidance.

Version 0.8.0, December 2003. Title changed to "Registering Web Map Services in *The National Map* Catalog." Major rewrite to sharpen the scope of the document, improve consistency with other reports, and make more consistent with current policies. Although much of the material is not changed, the editorial emphasis is shifted toward registration of WMSs in the catalog and away from partnership formation. Some material now covered in other documents removed.

Version 0.8.1, January 2004. Corrections made to URLs for OGC specification documents (OGC addresses have changed since early versions of this document). Other minor editorial corrections.

Version 0.8.2, January 2004. Changes to metadata statements to reflect 1/15/04 directive from the System Design Team about geospatial metadata implementation. The 1/15/04 directive added as a new attachment.

Version 0.8.5, February 2004. Ability to accept metadata from ArcIMS Metadata Services added. Explanation of a catalog requirement for geographic footprints added. Map sets [element sets] concept added.

Version 0.8.6, April 2004. Minor editorial errors corrected. URLs updated.

Version 0.8.7, June 2004. Significant editorial changes. Content updates to reflect current practices.

Version 0.8.8, September 2004. More information added to download section. Section 6 added with information about the service checker, system loads, GOS, and the GeoPartner Forum.

Version 0.9.0, November 2004. Significant editorial edits to improve readability and remove duplicated information. New section on data security added.

Version 0.9.1, March 2005. Minor editorial changes.

Table of Contents

Release Notes and Document History	i
Table of Contents	ii
1 Introduction.....	1
2 Definitions.....	2
3 Provider Responsibilities	3
4 USGS Responsibilities.....	4
4.1 Technical Support	4
4.2 Data Evaluation and Catalog Population	4
5 Overview of WMS Technical Characteristics	5
5.1 OGC WMS Compliance	5
5.1.1 Correct Support of GetCapabilities and GetMap Requests	5
5.1.2 ArcIMS Connector.....	5
5.2 Spatial Reference System (SRS).....	6
5.3 Data and Metadata Requirements	6
5.4 Data Download Implementation	6
6 Deliverables From Provider to CST	8
6.1 Summary of Provider/Service/Layer Information	8
6.1.1 Data Template.....	8
6.1.2 Complete Examples	8
6.2 Required Information.....	10
6.2.1 Service Address	10
6.2.2 Layer Metadata	12
6.2.3 Geographic Footprints	13
6.3 Recommended information.....	13
7 Other information.....	15
7.1 Security	15
7.2 Service Monitoring and the Catalog Service Checker	16
7.3 System Loads	18
7.4 Geospatial One-Stop	19
7.5 GeoPartner Forum.....	19
8 Points of Contact.....	20
9 References.....	20
Attachment A. Relationship of Provider Metadata to <i>The National Map Viewer</i>	21
Attachment B. Guidelines for Making New Data Visible	23
Attachment C. System Design Team Metadata Directive, 1/15/2004	25

1 Introduction

Web services are the technological basis for *The National Map*. The USGS is committed to supporting open standards and their implementations. *The National Map* will be implemented, to the fullest extent possible, using geographic information services (such as Web Map Services (WMS)) that conform to Open Geospatial Consortium¹ (OGC) specifications.

Data for *The National Map* comes from many providers. Each provider runs one or more services, each of which may serve many data layers. Data providers can be either within the USGS (for example, national data programs), other government organizations at the Federal, State, or local levels, or private companies. The number of providers, services, and layers is not limited.

Information about provider organizations, the services they run, and the data these services provide is managed by *The National Map catalog*. The catalog database is essentially an inventory of WMS data sources that contribute to *The National Map*. Applications get information about data locations from the catalog, as opposed to each application maintaining its own inventory of data sources.

Applications access the catalog database through a service that conforms to an OGC discussion paper. Having such a service, accessed through a published interface, means that applications can be written to be instantly and automatically responsive to changes in data sources. As the number of data sources grows, the ability to respond to changes without altering each application becomes very important.

Populating and maintaining the catalog is the responsibility of catalog support seams (CST) in USGS production mapping centers. Such teams exist at Mid-Continent Mapping Center (MCMC) and Eastern Region Geography (ERG). Catalog maintenance responsibilities may be distributed more broadly, possibly even to non-USGS organizations, as systems and procedures mature.

Data providers are responsible for

1. Serving their data through OGC-compliant WMS².
2. Providing appropriate metadata about the organization, services, and geospatial data.
3. Transmitting metadata and necessary technical information to the CSTs.
4. Assuring the content, position, and attribute accuracy of geospatial data.

The Catalog Support Teams are responsible for

1. Providing appropriate technical assistance to providers.
2. Evaluating WMSs under consideration for use by *The National Map*.
3. Adding information necessary to create a national application. Examples include applying standard symbols, setting display scales, setting relative layer priorities, and correctly

¹ Formerly "Open GIS Consortium." This organization changed its name in August 2004. The acronym OGC was not affected by the name change. The older name is still contained in many documents, and will continue to appear in quotations and reference lists for some time.

² A WMS is one of a more general class of services, collectively called "Open Web Services" or "OGC Web Services" (OWS in both cases). WMS is used in this paper because it is the type of service of primary interest to *The National Map* in 2004. Feature services and other types of services are expected to become more important as the program grows.

associating data and metadata to give providers appropriate recognition in *The National Map* applications.

4. Continuing to develop standards and guidelines — both technical and procedural — to better define the relationship between *The National Map* and provider data sources.

The interaction between providers and CSTs may involve other USGS teams and personnel, such as Mapping Partnership Office (MPO) staffs, Coordination and Requirements (C&R) staff, other design teams, and different levels of management. In all cases, the objective is to make provider data available to *The National Map* applications. The overall process is referred to in this document as **registering** services in the catalog.

This document is a specialized technical paper under the broader plan titled "A Process for Adding Partner Contributions to *The National Map*," written by the Implementation Partnership Team. See the USGS Partnerships page at <http://acis.er.usgs.gov/partnerships/> for copies of both documents (this page can be accessed only from the usgs.gov domain).

The latest version of this document can also be retrieved from <http://thor-f5.er.usgs.gov/nmcatalog/>. This page can be accessed by anyone, but is not linked from any other public USGS pages.

2 Definitions

- **Open Geospatial Consortium Web Map Services.** This phrase is variously shortened to **OGC-compliant WMS**, **OGC WMS**, or simply **WMS**. Reference [2] says "...most Web mapping applications today are still inseparably tied to a specific server implementation. In other words, the Web client is hard-coded to interact with a particular vendor's proprietary map server implementation...to address this problem, the OGC developed a non-proprietary Web mapping approach based on open interfaces..."

The use of OGC-compliant services is consistent with the "Implementation Plan for The National Map" (8/15/03 draft), which states: "The USGS intends to use the OGC specifications whenever possible...At this time, the only specification that the USGS has endorsed is the Web Mapping Service (WMS,version 1.1.1)."

Systems for delivering maps over the Internet are offered by many software vendors. Such systems go by a variety of similar-sounding names, such as "web mapping systems" and "online mapping services," in addition to proprietary names such as "MapObjects Internet Map Server" or "GeoMedia WebMap." In this document the term **Web Map Service (WMS)** will be used exclusively to refer to the OGC 01-068r2 specification [3] and its implementations.

- **The National Map catalog**, or simply the catalog. A database that inventories and tracks the owner organizations, services, and layers of WMSs that contribute to *The National Map*. This database is maintained centrally by the USGS and is populated primarily with metadata provided by data providers. The catalog database has an associated information service [8,9] that allows applications to access catalog data through a standard interface. Applications are therefore not required to understand the internal organization of the database.
- The **Phase D viewer** or **public viewer** is currently the primary application of *The National Map*. It is not the only application or even the only viewer, but it is currently the only public,

general-purpose, graphical application of *The National Map* that gets information about WMSs from the catalog service.

- **Data provider.** The owner of a WMS that serves data to *The National Map*, or the MPO/GSO that represents the actual data owner. Providers can be programs within the USGS or partners outside the USGS. This report discusses information that providers must transmit to the CST. In some cases, such as when a WMS is owned by a USGS national data program, communication will be direct from the provider to the CST. In other cases, such as a State or county data partner, these communications will be through a C&R or MPO/GSO representative.
- **Metadata.** Three types of metadata are relevant to this discussion:
 - **Organization or provider metadata** refers to information about the organization that owns or serves data. In *The National Map*, provider metadata is implemented by associating one URL with each provider organization. This URL typically points to the organization's home web page.
 - **Service metadata** is information about an information service. At the present time, Web Map Services are of primary interest. Service metadata is delivered by the service through XML documents as specified by the OGC WMS specification. Service metadata is extremely important to the smooth operation of *The National Map* applications, but is of little interest to most end users. Consequently, *The National Map* public viewer does not display service metadata.
 - **Layer metadata** is a synonym for geospatial metadata, cartographic metadata, or geographic metadata³. In *The National Map* implementation, each layer must have exactly one metadata URL, which points to a file or metadata service. Layer metadata is displayed by *The National Map* public viewer.
- **Catalog Support Team (CST).** Teams in USGS mapping centers responsible for data entry and maintenance of the catalog database. These teams work with providers and MPOs to add services to *The National Map*.
- **Mapping Partnership Office (MPO), Geographic Science Office (GSO), Coordination and Requirements (C&R), and liaison** are all synonyms, and all refer to people and offices that are responsible for negotiating partnerships with non-USGS organizations.

3 Provider Responsibilities

- Providing data to *The National Map* indicates a willingness to share geospatial data by publishing that data in Web services. Web services are inherently discoverable, and a major objective of *The National Map* is to make participating services **more** easily discoverable. The issue of data security in Web services is complex, and is briefly discussed in section 7.1 of this document.
- The fundamental technical requirement of providers is to deliver data through an OGC-compliant Web Map Service. See section 5 for a discussion and references.

³ In this document only. These terms may not be synonyms in other contexts.

- The provider's WMS must be configured to support the EPSG⁴ spatial reference system 4326 – WGS 84, in units of decimal degrees. The WMS may support any other SRSs desired by the data producer.
- Registering data in *The National Map* requires discussion and collaboration between the USGS and the WMS owners. Each source WMS must meet certain technical requirements, and the location and characteristics of the WMS must be communicated to the CST. These requirements cannot normally be met with a one-time exchange of information, but require an iterative, ongoing discussion between the USGS and the data provider.

4 USGS Responsibilities

4.1 Technical Support

For potential providers that are new to WMS technology, the CST will provide some assistance to set up services. Assistance with installing the ArcIMS connector has been particularly helpful to several partners in the past. The CST can also assist with altering map service configuration files to make them more consistent with *The National Map* applications.

This offer is not meant to imply that CST or other USGS personnel are unusually knowledgeable about WMS technology. In some instances, the data provider may be farther along the learning curve than we are. In all cases, discussions about technical issues between the two organizations will be helpful and probably necessary.

The USGS will also provide technical assistance with geospatial metadata creation on request.

4.2 Data Evaluation and Catalog Population

A major objective of *The National Map* is to use data from many WMSs in one application to create a nation-wide virtual map. This map includes many scales, many data layers, is reasonably seamless, consistent, and somewhat customizable. The mechanism for creating this map is for the USGS to maintain an inventory database – the **catalog** – of WMSs that contribute to *The National Map*. Given the WMS address and metadata information described in section 6.2, parts of the Catalog can be automatically populated using information from GetCapabilities queries. However, achieving the cartographic or geographic consistency desired for *The National Map* requires some manual inspection and evaluation of data.

These evaluations are a type of quality assurance of *The National Map* data. In general, the USGS does not own the spatial objects of data in *The National Map*, so this is a different kind of quality assurance than practiced in traditional cartography. A discussion of *The National Map* inspection and quality assurance practices is presented in reference [10].

⁴ European Petroleum Survey Group. See <http://www.epsg.org/>

5 Overview of WMS Technical Characteristics

5.1 OGC WMS Compliance

The fundamental requirement of providers is to deliver data through a WMS⁵. At the highest level of abstraction, this means delivering the following via HTTP upon request (from reference [1]):

1. User-selected spatial subsets through the GetMap request.
2. Service metadata in an XML file through the GetCapabilities request. Note this is metadata for the map service, not metadata about either the provider organization or the GIS data.

An optional third capability of a WMS is to deliver information about features in response to the GetFeatureInfo request.

In the event the above requests cannot be satisfied, appropriate error messages must be returned in an XML file.

The precise meanings of these requirements are defined in reference [3], and explained at varying levels of detail in references [1] and [2].

5.1.1 Correct Support of GetCapabilities and GetMap Requests

The WMS Implementation Specification [3] says the response to a GetCapabilities request shall be in the Extensible Markup Language (XML) [XML 1.0] and shall be valid according to the XML Document Type Definition (DTD)⁶. This requirement must be interpreted rigorously. The GetCapabilities response is parsed by USGS software to obtain a variety of information about the map service and its GIS data layers. These processes are called **harvesting**⁷ data from WMSs into *The National Map*. If the GetCapabilities response contains XML syntax errors, or is missing information required by the WMS specification, the harvesting will fail.

The response to a GetMap request must also conform rigorously to the WMS specification, but in practice this is less of a problem. GetMap responses, unlike XML documents, cannot be created or edited by hand, so it is unlikely that a service that does not construct proper GetMap responses will work at all.

5.1.2 ArcIMS Connector

Most data providers are "...tied to a specific server implementation [2]," usually, but not always, ArcIMS. However, this is not as big a problem as it might first appear. Reference [1] says "If you are already providing [spatial data] from a web server, then WMS-compliance is relatively simple..."

ArcIMS "out of the box" uses a non-OGC interface based on ArcXML. ESRI provides software called the **ArcIMS OGC WMS Connector** that allows ArcIMS to provide services that adhere to the OGC WMS specification. Reference [2] includes a cookbook tutorial on this connector.

⁵ As mentioned earlier, "WMS" is one instance of the class of Open Web Services (OWS). In the not-too-distant future, the WMS requirement stated here will change to a more general OWS requirement. But we expect to focus mostly on the WMS case through at least FY04.

⁶ See Annex A.1 of reference [3] for details. Another good source of WMS XML information is <http://www.digitalearth.gov/wmt/xml/>

⁷ Using "harvest" to refer to automated capture of service capabilities is not original. However, we have not found a formal definition.

An important point is that the data provider does not have to choose between the ArcXML interface and the OGC interface. The provider's map service can use **both** interfaces if the ArcIMS OGC connector is installed. Requests for data sent to the map server tell it which interface to use for responses. Existing non-OGC applications that run against a provider's map service therefore need not be affected by *The National Map*.

Other GIS vendors offer similar software to make their systems OGC-compliant. Map service software designed from the ground up to be OGC-compliant also exists; the most significant example at this time is probably the open-source MapServer from the University of Minnesota.

5.2 Spatial Reference System (SRS)

The WMS must support the WGS 84 geographic coordinate system; that is, a call to GetCapabilities must return EPSG⁸ code 4326 for the SRS. The WMS may support any other SRSs desired by the data producer. The USGS is analyzing projection options, and may allow other SRS options or have additional requirements for *The National Map* in the future.

5.3 Data and Metadata Requirements

Registering data in *The National Map* requires discussion and collaboration between the USGS and the WMS owners. Each source WMS must meet certain technical requirements, and the location and characteristics of the WMS must be communicated to the CST. These requirements are described in detail in section 6.

Internet data registries and clearinghouses are becoming common, and it is a USGS policy objective to promote one-stop registration of geospatial data. In particular, we do not want partner organizations to be required to register with both *The National Map* and Geospatial One-Stop (GOS). Implementation of this limited version of one-stop registration is underway. At this writing, the following are true:

- Registration with *The National Map* is adequate to insure registration with GOS. There may be a lag of a few days between when data becomes visible in *The National Map* and when it becomes visible in GOS.
- Registration in GOS does not yet lead to automatic registration in *The National Map*. This interface is in work, and should be partially implemented sometime in the summer or early fall of 2004. Because of fundamental differences in *The National Map* and GOS systems, registration in this direction will never be fully automatic. In particular, *The National Map* seeks to present consistent and seamless data, which requires more metadata. Actual duplication of data entry will eventually be eliminated, but GOS registration will never be completely adequate for *The National Map* registration.

5.4 Data Download Implementation

A WMS serves a picture of a map. If vector GIS data was used to construct this picture, the coordinate values and feature attributes of those data are not preserved in the response to a GetMap request.

The National Map intends to provide GIS data download capabilities that do preserve this information, but such capabilities are outside the scope of the WMS specification. The

⁸ European Petroleum Survey Group. See <http://www.epsg.org/>

specifications that define these capabilities are still being developed by OGC, and are not yet widely implemented in commercial software. Download capabilities will therefore be implemented in *The National Map* in phases over the next several years.

Two OGC specifications, related to but independent of the WMS specification, are relevant to this issue:

- Web Feature Services [4] return discrete geospatial features with their coordinate geometry and attributes. Version 1.0.0 of the WFS specification has been approved, but is not yet widely implemented in commercial software. ESRI, however, does have a WFS connector available from their download site.
- Web Coverage Services [5] return collections of geospatial features and data. The specification says "WCS provides access to potentially detailed and rich sets of geospatial information, in forms that are useful for client-side rendering, multi-valued coverages, and input into scientific models and other clients." Like the WFS, the WCS specification is not yet widely implemented.

It will obviously take time for these specifications and their implementations to mature. In the interim, *The National Map* viewer project has implemented non-OGC data download using ArcIMS download capabilities.

A commitment to work together to implement data download can therefore be included in agreements when both of these criteria are satisfied:

- ArcIMS is the service platform of the WMS.
- The partner's WMS is configured to permit data download.

If the partner agrees to allow file download, the issue of file names becomes important. ArcIMS download is accomplished with the GET_EXTRACT request, which returns a Shapefile. ArcIMS provides two methods for naming the extracted Shapefile. The default method is to use the layer's *id* attribute value. This is usually a number, so filenames are things like 2.shp, 3.shp... Such names are obviously neither descriptive nor unique.

We therefore request that providers use the second method for naming extracted Shapefiles, which is to specify a name by setting the OUTPUTFILE parameter in the ArcIMS configuration file. We suggest that file names have the form *servicename_layername.shp*. Including both the service name and the layer name in the Shapefile name may not absolutely guarantee uniqueness, but is a relatively straight-forward way to make the filename both descriptive and reasonably unique.

The USGS has no right to demand that partners conform to this, or any other, convention. But we ask that everyone concerned be aware that as *The National Map* grows, the proliferation of downloadable files with simple integer names will become increasingly confusing, and make all our organizations look unprofessional.

For technical information on the ArcIMS GET_EXTRACT command, see the "ArcXML Programmer's Reference Guide." An online copy of the relevant section can be found at http://downloads.esri.com/support/documentation/ims_/ArcXML9/Support_files/elements/using_get_extract.htm

6 Deliverables From Provider to CST

This section contains a fairly long list of things that must be known, or at least discussed, to include a new service in *The National Map*. Most of these things are not technical requirements, but rather information needed to make data display appropriately, or appear properly integrated, or give the data owner proper credit, or track the data source. These details are normally exposed through ongoing dialog between the USGS and the data provider. It is **not** a requirement for the provider to gather all this information together in one place "up front."

Section 6.1 summarizes data needed to register a new service or layer in the catalog. Section 6.2 describes each data element in more detail.

6.1 Summary of Provider/Service/Layer Information

The following two sections illustrate the information needed to add a service and its layers to the catalog. Section 6.1.1 is a generic description. Section 6.1.2 provides two examples based on existing partnerships.

6.1.1 Data Template

In the following outline, elements in **bold** type are required for either technical or policy reasons. Other elements are in some sense optional, but this should not be taken to mean they are unimportant.

Provider name

Provider metadata URL

Technical point of contact

Footprint, if appropriate (see 6.2.3)

For each service owned by the provider

Service name

Service address

Service viewer status

OGC version

For each layer within the service

Layer name

Layer label

Viewer status

Footprint, if not applied at the service level

Nominal scale

Viewscale range

Geospatial metadata URL

Legend URL

Theme and subtheme [may be repeated; a layer can be associated with more than one theme]

6.1.2 Complete Examples

The following two examples are services already registered in the catalog.

Example 1. BTS roads

Provider: BTS

Provider metadata: <http://www.bts.gov/about/>

Technical point of contact: [name, phone and email of someone in the provider organization]

Service Name: USGS BTS Roads WMS (EDC)
Service Type: WMS
Viewer Status: PUBLIC
OGC Version: 1.0.0
Footprint: all layers use the footprint of the contiguous 48 States
Service address:
 http://gisdata.usgs.net/servlet/com.esri.wms.Esrimap?ServiceName=USGS_WMS_BTS_Roads

Layer Name: Roads BTS
Layer Label: Roads BTS
Viewer Status: PUBLIC
Viewscale range: 0.00001 to 0.00118973048
Metadata URL: <http://www.gis.bts.gov/website/gdt/Dynamap1000.html>
Legend URL:
 http://gisdata.usgs.net/website/Map_Studio/legends/Legend_BTS.gif
Theme/subtheme: Transportation/Roads

Layer Name: Road Labels BTS
Layer Label: Road Labels BTS
Viewer Status: PUBLIC
Viewscale range: 0.00001 to 0.00014276765
Metadata URL: <http://www.gis.bts.gov/website/gdt/Dynamap1000.html>
Legend URL: no legend
Theme/subtheme: Geographic names/Names

Example 2. York County, South Carolina

Provider: York County
Provider metadata: <http://maps.yorkcountygov.com/>
Technical point of contact: *[name, phone and email of someone in the provider organization]*

Service Name: York County SC WMS (NC OneMap)
Service Type: WMS
Viewer Status: PUBLIC
OGC Version: 1.1.0
Footprint: all layers have extent defined by boundaries of York County, SC
Service address:
maps.yorkcountygov.com/servlet/com.esri.wms.Esrimap?ServiceName=usgs_urban

Layer Name: city
Layer Label: York Co. City (NC OneMap)
Viewer Status: PUBLIC
Viewscale range: 0 to 0.0012
Metadata URL:
<http://maps.yorkcountygov.com/gisonline/metadata/City%20Boundary.htm>
Legend URL: no legend
Theme/subtheme: boundaries/incorporated place

Layer Name: ortho200
Layer Label: York Co. Aerial Photography -200 ft (NC OneMap)
Viewer Status: PUBLIC
Viewscale range: 0 to 0.000214151486530036
Metadata URL:
<http://maps.yorkcountygov.com/gisonline/metadata/2000%20Orthophoto%20400-scale.htm>

Legend URL: no legend
Theme/subtheme: orthoimagery/other imagery

Layer Name: ortho400
Layer Label: York Co. Aerial Photography - 400 ft (NC OneMap)
Viewer Status: PUBLIC
Viewscale range: 0 to 0.000214151486530036
Metadata URL:
<http://maps.yorkcountygov.com/gisonline/metadata/2000%20Orthophoto%20400-scale.htm>
Legend URL: no legend
Theme/subtheme: orthoimagery/other imagery

Layer Name: parcels
Layer Label: York Co. Parcels (NC OneMap)
Viewer Status: PUBLIC
Viewscale range: 0 to 0.000214151486530036
Metadata URL:
<http://maps.yorkcountygov.com/gisonline/metadata/Land%20Parcels.htm>
Legend URL: no legend
Theme/subtheme: other/parcels

Layer Name: roadsarc
Layer Label: York Co. Roads (NC OneMap)
Viewer Status: PUBLIC
Viewscale range: 0 to 0.000356919144216726
Metadata URL: <http://maps.yorkcountygov.com/gisonline/metadata/Roads.htm>
Legend URL: no legend
Theme/subtheme: transportation/roads

Layer Name: roadslbl
Layer Label: York Co. Road Labels (NC OneMap)
Viewer Status: PUBLIC
Viewscale range: 0 to 0.000356919144216726
Metadata URL: none
Legend URL: no legend
Theme/subtheme: transportation/misc. transportation

6.2 Required Information

6.2.1 Service Address

The most technically important piece of information is the URL address(es) of the OGC service(s) that will serve data to *The National Map*. Because this information is so important, a detailed explanation of the service URL is given here, primarily for the benefit of USGS MPO employees.

The service address URL is **not** the same thing as the URL to the data provider's own home page or viewer application. While the service address can often be reverse-engineered from these other addresses, it helps avoid misunderstandings if the service address is provided explicitly.

The following service URL is an example from the Sedgwick County Kansas partnership, and is typical of services implemented in ArcIMS. Variations and the general form are discussed below.

`http://gis.sedgwick.gov:80/servlet/com.esri.wms.Esrimap?ServiceName=sedgwick_tnm`

where

- `gis.sedgwick.gov:80` is the server hostname with optional port number
- `servlet` is a directory path to a piece of software
- `com.esri.wms.Esrimap` is the name of some software that processes the requests
- `?` separates the software name from the parameter list
- `ServiceName=sedgwick_tnm` is what the service calls itself. This is one instance of a query parameter.

The path and software in this example – `/servlet/com.esri.wms.Esrimap` – is the ArcIMS default. Use of the default is common, but is neither universal nor required. It is also possible for the service and its OGC connector to have different paths; in this case, both paths must be given to the CST.

The parameter `ServiceName`, while common, is not required by the OGC WMS specification. The path can contain complete information about the service location, making the `ServiceName` parameter unnecessary. Section 6.2.2 of the WMS specification gives the general form of an OGC Web Service Request as

`http://host[:port]/path?{name[=value]&}`

`[]` = 0 or 1 occurrences

`{ }` = 0 or more occurrences

The Sedgwick county example above is a legal URL, but if typed into the address bar of a web browser it will return an error message. Though a legal URL, it is not the address of an HTML file, nor is it a well-formed request to a Web service. It is simply the *location* of a Web service.

Adding more parameters to the request makes the string a well-formed request to an OGC Map Service. Given the host, path, and service name, a parameter list such as the following can be entered into the address bar of a browser:

**`http://gis.sedgwick.gov:80/servlet/com.esri.wms.Esrimap?
ServiceName=sedgwick_tnm&request=GetCapabilities&service=WMS`**

(A line break and indent are added here to improve readability, but the request must actually be one uninterrupted string).

Where

- Parameters are separated by `&`
- `request=GetCapabilities` is one of the three OGC-defined queries to a WMS. `GetCapabilities` is essentially a request to the service to describe itself.
- `service=WMS` specifies that the query is to a Web Map Service, as opposed to (say) a Web Feature Service.

This request should return an XML document that displays in the browser. The XML file, though designed primarily for automated parsing and data extraction, is human-readable. Of particular interest is the list of descriptions for all the layers served by this particular service. These entries have a form similar to:

```
- <Layer queryable="0" opaque="0" noSubsets="0">
  <Name>County Boundary</Name>
  <Title>County Boundary</Title>
  <SRS>EPSG:4326</SRS>
  <LatLonBoundingBox minx="-97.81041" miny="37.468502" maxx="-
    97.14536" maxy="37.918404" />
</Layer>
```

This file can be inspected to confirm that the list of layers looks reasonable and complete. It should be consistent with information transmitted in documents about what layers are part of the service. All layers should include the "EPSG:4326" string, which says the service can provide data in WGS84 geographic coordinates.

6.2.2 Layer Metadata

The National Map requires FGDC-compliant metadata for each layer. At this writing, there are two acceptable mechanisms for providing these data to the CST:

- Each layer can have an associated FGDC metadata file. This file must be displayable in a Web browser, and must therefore be in HTML, plaintext, XML, or some other browser-displayable format. The USGS prefers not to have custody of the actual physical metadata files. It is preferable to keep the data and metadata together under the control of the data owner.
- Metadata can be served through a metadata service, such as ESRI's ArcIMS Metadata Service. The service dynamically generates temporary metadata files, which can be referenced as a URL. To access this type of service one needs to know:
 - The **DocumentID** (an id that is unique within the service). "docID" in the example below.
 - The **MetadataServiceURL**. "serviceURL" in the example below.
 - The **ServiceType** (to distinguish between (for example) an ESRI metadata service and an FGDC clearinghouse service). "serviceType" in the example below.

This information must be provided to the CST if metadata are to be obtained from a service. It is stored in the catalog as one URL string. For example:

```
http://mcmc.er.usgs.gov:8080/metadata_app?
serviceURL='http://datamil.udel.edu/servlet/com.esri.esrimap.Esrimap?
ServiceName=DelawareMetadata'&
docID={E219101A-E1BE-49FE-91E3-B64CE428AAC9}&
serviceType=ESRI_Metadata
```

(Linebreaks are inserted in this example to improve readability.)

Layer metadata is not a technical requirement for registering a service in the catalog, it is a procedural requirement expressed in a directive from the USGS System Design Team in January 2004 (see Attachment C). Services may be harvested into the catalog for evaluation or internal use without FGDC metadata, but will not be made publicly visible in *The National Map* applications without links to geospatial metadata.

Note that the provider must explicitly associate each layer with a metadata instance. Providing a list of layers and a list of metadata files, without relating the two lists to each other, is not adequate.

6.2.3 Geographic Footprints

Each layer is associated with exactly one geographic footprint. The footprints are used by *The National Map* graphic applications for guidance on when layers should be available for viewing, and for calculations to determine priority and ordering. Footprints are an unusual case of spatial coordinate data actually being stored in the catalog (in Oracle spatial objects) instead of being served from a WMS. The CST therefore needs a static GIS data file of a layer's geographic extent.

The normal case is for all layers within one service to use the same footprint, but this is not required. It is also possible for several services to use the same footprint; for example, national-coverage layers in different services use the same 48-State footprint.

Data providers **need not** provide footprint spatial data if their services and layers have National, State, or County extent (however, the fact that a layer has extent of (say) Alabama must still be transmitted). In these three cases, standard USGS datasets are used for footprints.

If the provider layers have some other footprint (a city corporate boundary, a State park, a National Forest...), then the data provider must give us a GIS coverage of the footprint boundary. Shapefile format is preferred. If the footprint corresponds to some well-defined polygon that is part of the data served by the service, the CST can create the footprint from these data instead of asking the partner to create a separate file.

6.3 Recommended information

WMSs can be registered in the catalog with nothing more than the service address(es) and geographic footprints described in section 6.2. Layer-level metadata is an additional USGS requirement for public data. However, providing only this minimal level of information may force the CST to make decisions that should be made by the data provider, or skirt important operational issues, or both. The data described below are therefore "not required" only in a narrow technical sense.

1. Organization metadata. If the provider wants credit for their contributions to *The National Map* in the viewer display, the CST must have a URL address for information about the provider. This is typically the home page of a data program or partner organization, but can be anything the provider wants it to be (provided the URL is legal and actually returns something).
2. Public vs non-public layers. The CST must be informed of any layers that should not be visible in *The National Map* public viewer. Reasons for making layers non-public include (but are not limited to):
 - Decisions by base data design teams. For example, the Hydrography program has directed that only NHD hydro data be public, except under very unusual circumstances.
 - Data provider preferences. Non-USGS providers may have layers they consider sensitive, and do not want displayed in USGS applications.
 - USGS-defined scope of *The National Map*. In some cases, a layer may not appropriately fit within any theme.

The default procedure is to mark all layers as REVIEW, meaning they can be seen in internal applications, but not in *The National Map* public viewer. Layers are switched to PUBLIC at the direction of the relevant design team or MPO, if they meet the other requirements specified in this document (e.g., they have layer metadata).

3. Layer scale. There are two kinds of scale:

- Scale class, or nominal data scale. For example, 1:24,000. This is not a precise concept, but is still useful. It can be thought of as the scale that would be appropriate for a paper plot of the layer.
- Viewscale. Scales in *The National Map* are expressed in **geographic degrees per screen pixel**. The catalog contains two such scales for each layer. These are hints to applications about the smallest scale and largest scale that are appropriate for display of the layer. For example, the GTOPO60 shaded relief elevation data is appropriate for display only at very small map scales. See reference [7] for more detail.

4. Legend URL. A long-term goal of *The National Map* is to apply consistent symbols through the mechanism of styled layer descriptors (SLD). Until SLDs are fully implemented, *The National Map* displays data using symbols supplied by the source WMS. The viewers can display a legend for these symbols. If such a legend is desired, the provider must create it as a .png image file with a transparent background and make it available through a URL.

5. Theme association. Every layer that is marked PUBLIC for display in *The National Map* must be associated with at least one data theme. For each theme the layer is associated with, the layer must further be associated with exactly one subtheme.

The appropriate base data design team should always be informed about layers added to their theme.

Non-PUBLIC layers that do not relate well to any theme can be assigned to the "other" theme.

If no guidance is provided, CST members will use their own best judgment to assign layers to themes.

6. Layer labels. The descriptive layer labels displayed on the right side of *The National Map* viewer are text strings stored in the catalog. These are not necessarily the same as the layer names used within the WMS (as shown in the response to a GetCapabilities query). The WMSs own names are sometimes too technical, too long, too short, etc. The USGS therefore creates our own labels, designed to be appropriate in *The National Map* viewer context. Providers and MPO/GSOs are encouraged to provide appropriate label strings to the CST.

At this time there are no formal editorial guidelines for these labels. Many labels have the form "Content (source)." For example, "Airport (Sedgwick County)" or "Primary routes (NC OneMap)".

7. Element set membership. The concept of an **element set**⁹ was introduced into the catalog design in early 2004. An element set is an arbitrary collection of layers that have been grouped together for some application purpose. For example, the Agency X may wish to display an application that uses some, but not all, of *The National Map* layers. At the same time, *The National Map* may wish to use some, but not all, of the layers from Agency X WMSs. Two map sets can be defined that contain the layers of interest to each organization. The two sets overlap, but neither is a subset of the other. Catalog-driven applications such as *The National Map* viewer can use element set information to customize the data shown in a particular instance of the viewer. This is a very powerful concept. When combined with viewer

⁹ In this document, **element set**, **layer set**, and **map set** are used as synonyms.

bounding-box start-up options, it allows a high degree of customization from a single application code base.

The default element set is *The National Map* map set. The STATUS attribute still applies, so it is possible for a layer to be part of this set without being displayed in *The National Map* viewer. A layer can belong to any number of sets.

8. Layer priorities. Displaying too many layers of similar data at the same time adds clutter and confusion to the viewer application. One of the most easily visualized examples is road layers. In some areas, we have national, state, and county layers of roads. The problem is that smaller coverages tend to be more current and more detailed, so allowing display of all three at the same time may not be desirable. But *not* allowing display of all three at the same time creates "holes" in the data that can have confusing visual effects, especially at the boundaries of the holes and islands.

Though not all instances of such problems can currently be resolved, many can. Data providers are encouraged to evaluate the appearance of their layers in the overall context of *The National Map* viewer and make suggestions.

7 Other information

This section contains addition background information that experience has shown to be of interest to data providers and MPOs.

7.1 Security

The National Map system is based on Web service technology. Broadly speaking, such services are not intended to provide secure data transport, but rather to provide easy access to published information. "Web Services Essentials" [11] summarizes the characteristics of a Web service this way:

- Is available over the Internet
- Uses a standardized XML messaging system
- Is not tied to any one operating system or language
- Is self-describing via a common XML grammar
- ***Is discoverable via a simple find mechanism*** [emphasis added]

If you have information you want to keep secret, it may not be a good idea to publish it in a Web service. Services provide security in the sense of protecting the computers and databases that host the data, not so much in the sense of protecting the requested copies of the published data itself.

One purpose of the catalog is to create "a simple find mechanism" for geospatial Web services. Although we impose some restrictions now, and will implement some stronger ones in the future, information in the catalog is generally intended to be open and accessible. The Oracle database itself is not accessible to users outside USGS firewalls, but the database has a service interface that is specifically designed to publish information about the location and general characteristics of WMSs. This interface isn't intended for interactive human use, so there is a certain degree of "security through obscurity," but that isn't much to depend on.

Services are inherently discoverable, even if they aren't in any registry or clearinghouse. We routinely find services simply by Googling. Automated crawlers that specifically target GIS services may not be common yet, but they will be.

Furthermore, the catalog only contains metadata – applications get the GIS data directly from the providers' services. The catalog holds information about the existence and location of services, some other service-level metadata, and a little geospatial metadata. It does not hold spatial objects or their attributes. So if a provider wants (for example) to publish pictures of maps, but wants to deny users the ability to download the original spatial objects, the responsibility for implementing that restriction lies with the service owner. The catalog can't and doesn't enforce **anything** having to do with how a user can query a service. At most, it provides hints about what kinds of queries might be useful.

The catalog developers are working on stronger catalog-level security that will allow us to relate individual users to individual datasets. The catalog will then hold information of the form "user X is allowed access to catalog entries about service S and layers Y and Z, provided that the person who claims to be user X can give the correct password." The user name and password information will be kept secure using regular Oracle security mechanisms (which are actually quite good), and will not be exposed through the service interface. But this, by itself, still isn't terribly strong security. We can't do anything to prevent people from sharing passwords; and at some point in this process, information about the location of services will still be delivered across the Internet in a stream of plain text.

Strong data security, the kind that provides a high degree of certainty that a dataset will never be compromised or distributed to someone you don't approve of, requires the infrastructure described in the previous paragraph, plus at least 1) a serious attitude toward passwords and 2) strong encryption of data streams. Neither are especially common in GIS systems today.

7.2 Service Monitoring and the Catalog Service Checker

The USGS runs software that automatically checks the status of the services that contribute to *The National Map*. There are at least three reasons this monitoring is important:

1. In the short run, it is useful to know what services are down. End users often contact the USGS with questions of the form "I saw this data yesterday, and today it isn't there..." A log of service availability helps diagnose these situations.
2. Also in the short run, applications can avoid time-out delays by checking with the catalog to see if a service is up before sending requests to the service. For this to be a benefit, the checker data must be very fresh. For this reason, services are checked several times per hour.
3. In the long run, it is valuable to have metrics for the overall reliability of *The National Map*'s data sources, and to know which services are most and least reliable. The catalog web site serves several reports that summarize *The National Map* reliability as measured by data from the catalog service checker.

The catalog service checker is conceptually simple. For each public¹⁰ service registered in the database, the checker sends an OGC GetCapabilities request to the service. There are three possible outcomes of this action:

¹⁰ "Public" means the service has at least one layer that is visible in *The National Map* public viewer. The service checker does not query services that do not meet this condition.

1. A legal OGC capabilities document is returned. "Success" is logged, with a timestamp, in a table of the database and the checker continues to the next service.
2. Something other than a legal OGC capabilities document is returned. This is typically an error message from the service, the host OS, or some part of the Internet infrastructure. The checker waits 300 milliseconds and tries again (the delay is to prevent the target server from being hit with a rapid barrage of requests). The checker will make up to five attempts before giving up. If all five attempts are unsuccessful, "failure" is logged and the checker continues to the next service.
3. Nothing is returned within a reasonable time. This probably means the computer hosting the service is down, but there are other possible causes. After 180 seconds the checker logs "failure."

The checker cycles through the complete list of public services. After completing one cycle the checker waits 900 seconds (15 minutes), then starts over. Each service is queried about 3.5 times per hour on average. At any given time, the catalog information about which services are functioning properly is accurate to within about 17 minutes.

This is reasonably straight forward, and when the checker, the host, and everything between are functioning more-or-less normally, the checker activity usually goes unnoticed.

The checking activity is sometimes referred to as "pinging" the services. This terminology is acceptable as shorthand or metaphor, but is not technically correct. **Ping** is a Unix utility used by system administrators to see if a computer is operating and also to see if network connections are intact. Ping was written in a more innocent era of network computing, and is associated with several security holes. Most computers today (including all USGS computers) have ping disabled. **The catalog service checker does not use ping.** Rather, it sends a request for an OGC capabilities document and evaluates the response. This is technically very different from a ping and carries no more security risk than any other Web service request.

OGC capabilities documents are usually quite small, typically five to 20 kilobytes. Such small documents can be retrieved frequently – dozens to thousands of copies per minute, depending on the size of the host computer and the bandwidth of its network connection – without adversely impacting the performance of the host.

The catalog service checker queries **services**, not **hosts**. One host can manage many services. Furthermore, as explained above, services that are not responding properly may be queried up to five times per cycle. So under some relatively unusual conditions, a host could be queried dozens or hundreds of times per hour, not the nominal 3.5 times per hour quoted above. There are no known cases of these queries creating a system overload, but the relatively large number of identical queries in the host's web logs can cause concern if the system administrator is not aware of the nature of the queries.

The checker can be adjusted to change the interval between queries, and the interval can be customized for particular services. It is also possible to exclude specific services altogether. If a partner has legitimate reasons for wanting their services checked less frequently, or temporarily not at all, we will try to arrive at a mutually satisfactory arrangement. However, we think service monitoring is important, and also believe it is important to gather data points at an average rate of at least two per hour per service.

The monitoring is not 100% reliable. Three relatively common errors are:

- False success may be registered if an OGC connector is up and serving static capabilities documents¹¹, even though the IMS behind the connector is down and the service is not delivering map images. The behavior in this case is vendor-dependent; it is not defined by the WMS specification.
- False success is registered if the service is responding to requests with documents that are syntactically legal but have incorrect content.
- False failure is registered if some other part of the overall Internet system (routers, firewalls...) is preventing queries or their responses from being properly delivered.

The first of these three is probably the most common cause of incorrect results, but we have no data on the frequency of any of these types of problems. Anecdotal evidence is that they happen often enough to be annoying, but not often enough to constitute a major problem.

7.3 System Loads

A frequently asked question is: "How will participation in *The National Map* affect system loads on partners' computers?"

The USGS cannot answer this question, for several reasons:

- Except for the catalog service checker described above, *The National Map* systems do not themselves initiate requests to partner services¹². *The National Map* provides the information about the **existence and location** of services to end users or applications, which then initiate queries to services of interest.
- Demand for data tends to be driven by local forces that the USGS doesn't necessarily know about, and are probably too complex to model anyway.
- Most services can also be accessed through applications that have no relationship to *The National Map*. States and counties that host WMSs usually have their own local applications. Sometimes these applications are much more heavily used than *The National Map* interface into the service.
- System tools that measure the load on a particular service are generally not available to the USGS (nor should they be), but only to the system administrators of the organization that hosts the service.

It is certain that participation in *The National Map* will increase, at least slightly, the load on a service and its host. Quantitative predictions about this increase are impossible. Intelligent guesses, based on knowledge of the GIS user base in a particular area, can be made, but local organizations are better qualified than the USGS to make these predictions.

¹¹ Dynamic capabilities documents are created by the service at the time they are requested. This is not required by the WMS specification – it is acceptable to store the capabilities document statically on the host file system. In this case, an OGC connector can return a capabilities document without actually querying the underlying map service. We have considered basing the service checker on GetMap queries instead of GetCapabilities queries to avoid this problem, but have rejected that option for now.

¹² This may sound a bit disingenuous, since USGS-owned viewers do initiate GetMap requests to provider services. But they do so only in response to directions from human users. The query rate is therefore not predictable.

Partners can share their experiences through the GeoPartner Forum (see section 7.5 below). Swapping anecdotes may be the best way to develop an understanding of the effect of *The National Map* on system loads.

7.4 Geospatial One-Stop

Internet data registries and clearinghouses are becoming common, and it is a Federal policy objective to promote one-stop registration of geospatial data. In particular, we do not want partner organizations to be required to register separately with both *The National Map* and Geospatial One-Stop (GOS). Implementation of this limited version of one-stop registration is underway. At this writing, the following are true:

- Registration with *The National Map* is adequate to insure registration with GOS. The USGS periodically derives GOS registration files from *The National Map* catalog and makes these data available to GOS.
- Registration in GOS does not yet lead to automatic registration in *The National Map*. Because *The National Map* requires more registration metadata than GOS, registration in this direction will never be fully automatic. Actual duplication of data entry may eventually be eliminated, but GOS registration will never be completely adequate for *The National Map* registration.

Interfaces to promote one-stop registration do not address the issue of a partner organization that wants to participate in one of these efforts, but opposes participating in the other. Such cases do not appear to have been considered during formation of high-level Federal GIS policy.

7.5 GeoPartner Forum

The University of Missouri at Columbia is hosting a site for partners to discuss technical issues pertaining to hosting Web services in support of *The National Map*. The forum was announced to the MPO network in mid-September 2004. The purpose of forum is to provide a mechanism for partners to share information and discuss issues among themselves. The USGS will participate, but it is our hope that the forum will not be primarily an "ask USGS" site. The forum is **not** a site for vendors or the general public to make comments and ask questions. Anyone can read postings, but only registered users can create new postings, and registration is restricted.

The forum is moderated from Mid-Continent Mapping Center.

The site currently contains two forums: Announcements and Partner Discussion. The **Announcements** forum contains read-only announcements about events and general notes. The **Partner Discussion** forum allows reply and posts by registered users. Currently, all visitors to the site are allowed to view either forum.

The forum web site is <http://www.geopartner.org>.

8 Points of Contact

This document is maintained by the Catalog Support Team (CST) at Mid-Continent Mapping Center. The primary points of contact for this team are

- Larry Moore, 573-308-3661, lmoore@usgs.gov
- Debbie Cochran, 573-308-3894, dcochran@usgs.gov
- Larry Jontz, 573-308-3753, ljontz@usgs.gov

Mail sent to the following CST mailbox will be read regardless of who might be on vacation or travel:

- Internet address: USGScatalog@usgs.gov
- Lotus address: GS-N-MCMC Catalog Support Team

Eastern Region Geography has its own CST. The point of contact for this team is Art Eckerson (703-648-4580).

9 References

The catalog Web site at <http://mcmcweb.er.usgs.gov/catalog> is the general repository for technical information about the catalog database, service, and registering data in *The National Map*.

- [1] "Guide to Distributing Your Data Products Via WMS 1.1.1, A Tutorial for Data Providers" <http://oceanesis.jpl.nasa.gov/esipde/guide.html>
- [2] "OpenGIS Web Map Server Cookbook, version 1.0.0, OGC 03-050, <http://www.ogcnetwork.org/docs/03-050.pdf>
- [3] "Web Map Service Implementation Specification" version 1.1.1, OGC 01-068r2, <http://www.opengeospatial.org/docs/01-068r2.pdf>.
- [4] "Web Feature Service Implementation Specification" version 1.0.0, OGC 02-058, <http://www.opengeospatial.org/docs/02-058.pdf>
- [5] "Web Coverage Service (WCS), Version 1.0.0", OGC 03-065r6, <http://www.opengeospatial.org/docs/03-065r6.pdf>
- [6] "Availability of *The National Map* Web Map Services." Internal USGS technical discussion paper. http://mcmcweb.er.usgs.gov/catalog/tnm_catalog_page2.html#techinfo
- [7] "Viewscales and Their Effect on The National Map Displays." Internal USGS technical discussion paper. http://mcmcweb.er.usgs.gov/catalog/tnm_catalog_page2.html#techinfo
- [8] "OGC Web Services Stateless Catalog Profile" OGC-IP Draft Candidate Specification (01-062) <http://thor-f5.er.usgs.gov/nmcatalog/2001-062StatelessCatalogProfile.pdf>
- [9] "*The National Map* Catalog Service: A Guide for Application Developers" http://thor-f5.er.usgs.gov/nmcatalog/catalog_api.pdf
- [10] "Quality Assurance of Cartographic Data in Catalog Operations, FY2004" http://thor-f5.er.usgs.gov/nmcatalog/quality_assurance.pdf
- [11] Cerami, Ethan, "Web Service Essentials." O'Reilly, 2002.

Attachment A. Relationship of Provider Metadata to *The National Map Viewer*

Figure A-1 is a screen shot of a portion of *The National Map* viewer. The area shown is in Shreveport, LA. The Louisiana Regional Applications Center (RAC) is the data partner providing much of the large-scale data for this area.

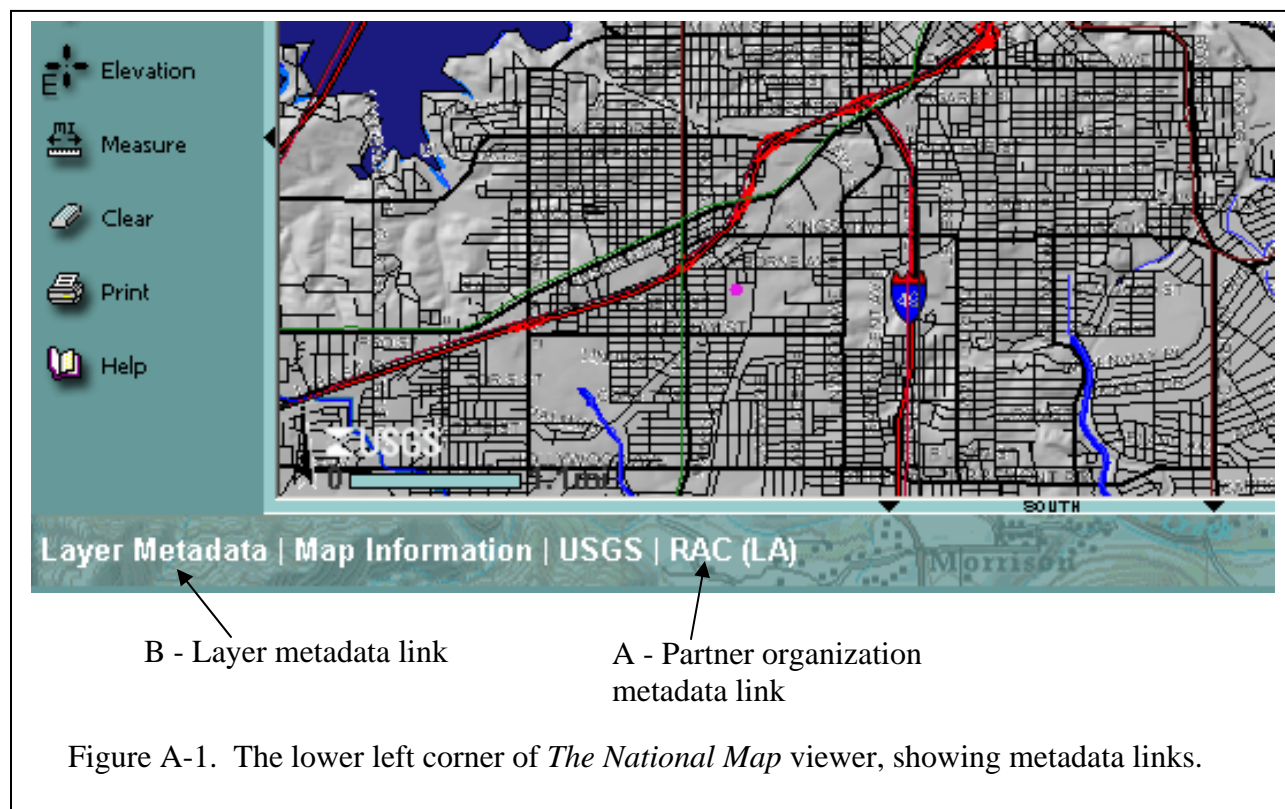


Figure A-1. The lower left corner of *The National Map* viewer, showing metadata links.

The hyperlink labeled "A - Partner organization metadata link" in Figure A-1 connects to a page titled "The National Map of Louisiana." This page is owned and maintained by RAC. It can be thought of as a different WMS application that uses some of the same data as *The National Map*. The link to RAC is displayed in the viewer because the catalog contains the information that RAC is the appropriate data partner in this area at this scale.

The link labeled "B - Layer metadata link" in Figure A-1 opens a window that displays the table shown in Figure A-2. The table shows all layers that are currently active, with a link to the layer's metadata, if any. The URLs for these links are provided by the data provider as described in section 6.2.

As directed by the System Design Team on January 15 2004, all layers are now required to have links to FGDC metadata.

The URLs are provided by the data provider as described in section 6.2.2, and the pages pointed to by the URLs are hosted and maintained by the data provider. These metadata links can contain data distribution and sales information. This is especially important in cases where the partner

prefers not to allow data download through *The National Map* applications. In these cases, the links back to the data producer will be the only way for end users to access the GIS data.



Layer Metadata

Min-Lon: -93.7891 **Max-Lon:** -93.7442
Min-Lat: 32.4835 **Max-Lat:** 32.5113

Projection: Geographic (NAD83)

Layer	Description	Metadata
FS Ranger District Boundaries	This dataset contains National Forest Ranger District boundaries for the lower 48 states and Alaska, including Puerto Rico.	Metadata
FS Wilderness Boundaries(Outline)	This dataset contains all Special Designated Areas (SDAs) for the lower 48 states, including Puerto Rico.	Metadata
FS Wilderness Boundaries(Polygon Fill)	This dataset contains all Special Designated Areas (SDAs) for the lower 48 states, including Puerto Rico.	Metadata
Forest Service Administrative Boundaries(Outline)	This dataset contains National Forest boundaries for the lower 48 states and Alaska ,including Puerto Rico	Metadata
Forest Service Administrative Boundaries(Polygon Fill)	This dataset contains National Forest boundaries for the lower 48 states and Alaska ,including Puerto Rico	Metadata
County Boundaries (USGS)	U.S. Counties (Detailed) represents the counties of the United States in the fifty states and the District of Columbia	Metadata
Parishes (LA)	This is a region dataset depicting the polygon boundaries of the 64 parishes comprising the state of Louisiana.	Metadata
Boundary (LA)	The dataset defines the state 'territorial' boundary of Louisiana.	Metadata
US NED Shaded Relief	The USGS National Elevation Dataset (NED) has been developed by merging the highest-resolution, best-quality elevation data available across the United States into a seamless raster format. NED is the result of the maturation of the USGS effort to provide 1:24,000-scale Digital Elevation Model (DEM) data for the conterminous US and	Metadata

Figure A-2. Active layers and metadata links.

Attachment B. Guidelines for Making New Data Visible

Technical information exchange is not the only communication needed to implement partnerships. Policy-making processes are outside the scope of this document, but a brief discussion is included here at the request of the CTM Steering Team, pending development of more complete guidelines by other teams.

The issue is how to coordinate the activities of several entities to implement partnership agreements. The three groups of primary interest are:

1. State liaisons in Mapping Partnership Offices (MPO) and the State and local organizations they work with. MPOs are responsible for negotiating agreements that conform to Discipline policies and contribute to *The National Map*.
2. Base data design teams are responsible for defining *The National Map* data content standards.
3. The Catalog Support Team (CST), under the direction of the Directory Design Team, is responsible for implementing decisions made by the other two groups.

All three groups are currently developing procedures within their own domains, but none of these activities are anywhere near complete. The CTM Steering Team has established the following general guidelines for implementing partnerships in the near-term:

- A. The MPOs have primary authority for making decisions about what non-USGS data are included in *The National Map*. These decisions should be informed by at least the following:
 - i. New data should contribute something new to *The National Map*, as opposed to being largely redundant with existing layers.
 - ii. New data should fit reasonably well into an existing theme. Failing this, there should be a design-team sponsor for creating a new theme(s) for the data. The MPO liaisons and their State counterparts are the judges of what is "reasonable," but other groups may react to these decisions (see D below).
 - iii. New data should conform to existing standards, content guidelines, and policy directives developed by design teams. Most data programs have not yet developed or distributed such standards. Where this is true, this guideline is equivalent to point A(ii) above.
- B. Base data design teams have the authority and responsibility to define standards and inclusion guidelines. These standards will eventually define the content of *The National Map* and therefore dictate the types of data that can be accepted in partnership agreements. However, defining these standards will be a long and evolutionary process. All groups should be aware that partnership negotiations are ongoing and cannot be suddenly disrupted by radical changes to content requirements.
- C. The CST is responsible for implementing partnership agreements by populating the catalog with partner service information. This involves many small decisions about (for example) layer priorities, display scales, and footprints. The CST will make most of these operational decisions, but will take guidance from the MPOs as appropriate. The top-level decision

about whether a layer should be included in or excluded from *The National Map* lies with the MPOs, not with the CST.

- D. All groups have both the right and the responsibility to monitor the results of this process. Use of internal software tools (such as SQL*Plus, MapView and MapConnect) that expose non-public details of the data organization is strongly recommended. Organizing WMS information in the catalog is a continuous process. Any of the three groups may request catalog changes to improve the coherence of the data. Major changes [for example, a decision to retroactively disallow a broad class of layers] will be discussed between the three groups, and possibly in other forums as well, before being implemented. Small changes will normally be implemented without discussion [for example, changing the scales at which a particular layer is made available for display].

Attachment C. System Design Team Metadata Directive, 1/15/2004

The following document was distributed by email on January 15, 2004 by the System Design Team. The May 14 deadline set in this memo was met; all layers of The National Map now have metadata links. However, many of these links still point to files that fail to meet the FGDC standard.

Assignments to Design Teams to Solve Metadata Problem

Problem: Only 38% of currently visible layers in *The National Map* catalog have metadata registered. Even our own holdings have only a 34% compliance rate. This is unacceptable. It's against our own policy, and has been a long-standing and growing embarrassment.

Background: The Standards and Directory Design teams have collaborated to collect the background material. The Standards Team reaffirmed that we have a clear policy already for metadata. This is an issue of compliance. The Directory Team summarized the current situation in a technical discussion paper "Geospatial Metadata in The National Map" (posted on internal partnerships web page). The entire issue was briefed to the IFORUM on Jan 12, 2004. Later that same day, the System Design Team discussed the issue and arrived at the following assignments to solve this problem:

Solution:

1. While we act to correct this problem, we need to prevent this situation from getting any worse. Therefore, effective **immediately**, we are asking the Catalog Support Team (a part of the Directory Design Team) or others trained and certified to make and maintain catalog entries to cease marking any content as publicly available unless it has URL-accessible metadata that meets the minimal FGDC and National Map metadata content standards.
2. The Standards Team is responsible for defining the minimum (normative) mandatory content for layer metadata.
3. We don't want to over-react by setting all non-compliant catalog content to "non-public" – this would confuse the public. However, we need to establish a reasonable timeframe to allow us to "catch-up" with this metadata population. Based on recommendations from the Standards Design Team who are familiar with what needs to be done, we are setting the due date for layer-level metadata population to May 14, 2004. Any layer not metadata compliant after this date will be marked "non-public" within the catalog.
4. Responsibility for populating: It is important to spread this workload to the groups who should have this as part of their normal job. This is not the responsibility of the standards or catalog teams – those groups can both provide advice and training, but the operational responsibility for layer-level metadata population (seeing that it gets done) belongs to different groups:
 - a. For any of the eight layers held and managed by USGS, the responsibility for populating layer metadata resides with the specific Base-Data (theme) Design Team.

- b. For any of the eight layers provided by external partners, the responsibility for populating layer metadata resides with the MPO or C&R staff that developed the partnership.
 - c. For USGS content other than the eight layers, the responsibility for populating layer metadata resides with the data steward for the particular data holdings.
 - d. For non-USGS content, metadata will be accessible by means of the GOS /USGS directory synchronization process assigned to the Directory Design Team to develop.
- 5. The Catalog Operations Team will provide reports to each group summarizing the layers that have missing layer-level metadata.
 - 6. To the extent this work is seen as new or unanticipated, the impact of this “catch-up” work should be reflected in the revised narratives being currently being prepared by each Center. If any group sees any major obstacles that would impact our goal of being 100% compliant with our own metadata policy by May 14, the System Design Team needs to be made aware of that compelling case as soon as possible.
 - 7. The principle here is making this part of the job. If training is required, there are many sources of training available. FGDC has several training programs, and our own standards people can also assist in any needed training.
 - 8. We also need to ensure that the technical connection between metadata generation engines employed by some of our internal holdings (NED and other SDDS holdings, for example) work well with the metadata functions in the viewer. We are asking the Leads of the viewer implementation and catalog teams to have technical discussions with Jean Paulson (for SDDS) and Steve Skelton (for GDA) to work out the automated interface specifications for generated layer metadata. This will be the mechanism we use to propagate source-file level metadata.

FAQ

If I can't easily get the information from a partner, can I simply populate metadata fields with “unknown”?

It depends on the particular field. Some fields permit an “unknown” value, while others do not. It is NOT an acceptable practice to simply value a field with the “unknown” value for the purpose of saving time or effort. There are certain mandatory fields that must be populated, and only certain fields can assume the value “unknown”. If you are unsure as to the minimum population rules, contact the Standards Design Team.

Where can I get information on what things I need to worry about when registering content for *The National Map*?

Specific technical requirement are described in the document “Registering Web Map Services in The National Map”, posted on the internal partnerships web site (See http://thor-f5.er.usgs.gov/nmcatalog/wms_register.pdf). Send questions and requests for assistance to “GS-N-MCMC Catalog Support Team”.

While the Standards Team is responsible for overarching standards (characteristics common to all themes; i.e., projections) and for some level of consistency, we have been deferring to the data theme leaders for content and accuracy standards.

The Standards Team is responsible for the metadata field population rules. Population rules specify which fields must be populated, and the domain of acceptable values. The Eight Base Data Design Teams are responsible for defining minimum content specifications (inclusion criteria), which would include tighter ranges of acceptable values for certain metadata fields.

Where exactly can I find the minimum metadata population rules?

FGDC metadata standard: <http://www.fgdc.gov/metadata/contstan.html>

The **workbook** for this standard: http://www.fgdc.gov/metadata/meta_workbook.html

We recommend using the workbook rather than the standard for help. It is easier to use and contains a graphic representation of each section of the standard.

USGS metadata standard: <http://rockyweb.cr.usgs.gov/nmpstds/metastds.html>

Although the USGS document is referred to as a standard, the Standards for the Preparation of Digital Geospatial Metadata is actually the USGS implementation of the FGDC standard. This 'standard' includes the elements from the FGDC standard that we decided to apply to our products, and has sample metadata for each data type and each scale that they were produced (DLG, DEM, DOQ, and DRG). These documents may be used as a guide for the proper elements to include, however, the content of the elements (i.e. the responses to the element) will need to be altered for each dataset.

It should be noted that the Standards Design Team is currently in the process of recasting our current product-oriented metadata standards to a form more applicable to the web-services model. However, the content in the current product-oriented standard will be sufficient until the services-oriented version can be completed.

How did you arrive at the May 14 deadline?

This is the recommendation from the Standards Team taking into account the following factors:

1. We want to correct this problem as soon as **practically** possible
2. The work to correct the problem is being spread over several Base Data teams and C&R/MPO offices
3. Some training will be required – this time is built in
4. Some time may be required to bring existing metadata into a URL-accessible form - this time is also built in
5. We are asking any group that can make a compelling case that the 14 May due date is unreasonable to make that case to the SDT as soon as possible. We don't want any surprises on May 15
6. We expect that communicating this decision/directive out to all affected teams should transpire NLT 1/14/04

Isn't this a policy issue?

No – we have an established policy (see Standards team ppt briefing). This is a compliance issue.

I thought the System Design Team was supposed to work in design issues, not operational issueswhat gives????

The Systems Design Team (and the individual Design Teams) have a dual responsibility – for design assignments, and for establishing operational procedures. The actual people that accomplish the work are in Centers and/or remote offices, but they are following the operational procedures defined by the design teams.

So, the actual assignment to fix this must ultimately come thru the Center Chiefs?

Yes – that’s why they will be cc’d in on this action request.

If this is part of the C&R job, it will take longer to add content – I thought we were to add as much content as we can as fast as we can.

Yes, this is part of the job, and if that means completing the **whole job** takes longer, so be it. There may have been a time when we skimped on metadata in favor of bulk content, but those days are gone (if they ever really consciously existed at all). The USGS has been a major supporter of metadata, and it simply sends the wrong message if we lack the commitment to follow-thru on this important aspect of geospatial data.

Is the problem that we simply don’t have any metadata, or that it exists, but the catalog doesn’t know about it?

We have a little bit of both, but from early reports, this appears to be mostly about getting the metadata we do have into a form the catalog can use. Generally, that means ensuring that the metadata (whether it’s static, or dynamically generated) resides at some URL.

What level of meta-data are we talking about here? Service, series, layer, product, file, or feature?

The immediate problem concerns layer-level metadata. Let’s solve that problem first. In *The National Map* implementation, each WMS layer should have exactly one metadata URL, which points to a file of metadata or to a metadata service. If a static file, the file must be in a browser-displayable format (HTML, plaintext, XML...).

We don’t have a problem with service metadata - our current catalog population process ensures that service metadata is populated and correct. Access to source file and/or product metadata is covered in action #8. As for feature-level metadata, we’ll attack that issue later. With the possible exception of NHD, very few of our holdings even have feature-level metadata. We will need to invest more in this area once we start moving into feature-oriented web services.